Procesamiento del Lenguaje Natural, Revista nº 59, septiembre de 2017, pp. 91-98 recibido 30-03-2017 revisado 16-05-2017 aceptado 22-05-2017

Enriching low resource Statistical Machine Translation using induced bilingual lexicons

Uso de lexicos bilingües inducidos para el enriquecimiento de un sistema de traducción automática estadística de pocos recursos

Han Jingyi, Núria Bel
Universitat Pompeu Fabra
Roc Boronat, 138, 08018, Barcelona
{jingyi.han, nuria.bel}@upf.edu

Abstract: In this work we present an experiment for enriching a Statistical Machine Translation (SMT) phrase table with automatically created bilingual word pairs. The bilingual lexicon is induced with a supervised classifier trained using a joint representation of word embeddings (WE) and Brown clusters (BC) of translation equivalent word pairs as features. The classifier reaches a 0.94 F-score and the MT experiment results show an improvement of up to +0.70 BLEU over a low resource Chinese-Spanish phrase-based SMT baseline, demonstrating that bad entries delivered by the classifier are well handled.
Keywords: Machine translation, phrase table expansion, bilingual lexicon induction, Natural language processing

Resumen: En este artículo presentamos un método para ampliar la tabla de frases de un traductor automático estadístico con entradas bilingües creadas automáticamente con un clasificador supervisado. El clasificador es entrenado con una representación vectorial en la que se concatenan el vector distribuido (Word Embeddings, WE) y una representación de agrupaciones de Brown (Brown clusters, BC) de 2 palabras equivalentes de traducción. El clasificador alcanza una F1 de 0,94 y el resultado de la evaluación del sistema de traducción automática entre chino y español muestra una mejora de hasta +0,70 BLEU, demostrando que las malas traducciones producidas por el clasificador son controladas bien por el sistema de traducción.
Palabras clave: Traducción automática, Expansión de vocabulario, Inducción de léxicos bilingües, Procesamiento del lenguaje natural

1 Introduction

Parallel corpora are one of the key resources that support Statistical Machine Translation (SMT) to learn translation correspondences at the level of words, phrases and treelets. Although nowadays parallel data are widely available for well-resourced language pairs such as English-Spanish and English-French, parallel corpora are still scarce or even do not exist for most other language pairs. The translation quality with no data suffers to the extent of making SMT unusable.

Many researches (Fung, 1995; Chiao and Zweigenbaum, 2002; Yu and Tsujii, 2009) attempt to alleviate the parallel data shortage problem by using comparable corpora which still are not readily available for many language pairs. Monolingual corpora, on contrary, are being created at an astonishing rate. Therefore, in this work, we propose to extend an SMT translation model by augmenting the phrase table with bilingual entries automatically learned out of non necessarily related monolingual corpora. The bilingual lexicon was delivered by a Support Vector Machine (SVM) classifier trained using a joint representation of word embedding and Brown cluster of translation equivalents as features.

The main contributions of this paper are: (1) We present a supervised approach to automatically generate bilingual lexicons out of unrelated monolingual corpora with only a

ISSN 1135-5948

small quantity of translation training examples. (2) We prove that enriching an SMT phrase table using all the results, including the errors delivered by the classifier, is indeed a simple and effective solution.

The rest of the paper is structured as follows: section 2 reports the previous works related to our approach; section 3 describes our supervised bilingual lexicon learning method; section 4 sets the experimental framework; section 5 reports our test results; and section 6 gives the final conclusion of the work.

2 Related work

The use of monolingual resources to enrich translation models has been proposed by different researches. For instance, (Turchi and Ehrmann, 2011; Mirkin et al., 2009; Marton, Callison-Burch, and Resnik, 2009) used morphological dictionaries and paraphrasing techniques to expand phrase tables with more inflected forms and lexical variants. Another line of work exploits graph propagation-based methods to generate new translations for unknown words. For instance, Razmara et al. (2013) proposed to induce lexicons by constructing a graph on source language monolingual text. Nodes that have related meanings were connected together and nodes for which they had translations in the phrase table were annotated with target side translations and their feature values. A graph propagation algorithm was then used to propagate translations from labeled nodes to unlabeled nodes. They obtained an increase of up to 0.46 BLEU compared to the French-English baseline. Similary, Saluja et al. (2014) presented a semi-supervised graph-based approach for generating new translation rules that leverages bilingual and monolingual data. However, all these methods generate new translation options by depending on existing knowledge of a baseline phrase table.

In order to create new entries, Irvine and Callison-Burch (2013) used a log-linear classifier trained on various signals of translation equivalence (e.g., contextual similarity, temporal similarity, orthographic similarity and topic similarity) to induce word translation pairs from monolingual corpora. Irvine and Callison-Burch (2014) used these induced resources to expand the SMT phrase table. Since much noise was introduced, 30 monolingually-derived signals needed to be applied as further translation table features to prune the new phrase pairs. Experiments were conducted on two different language pairs. An improvement of +1.10 BLEU for Spanish-English and +0.55 BLEU for Hindi-English was achieved.

The challenge of bilingual lexicon induction from monolingual data has been of long standing interest. The first work in this area by Rapp (1995) was based on the hypothesis that translation equivalents in two languages have similar distributional profiles or co-occurrence patterns. Following this idea, (Koehn and Knight, 2002; Haghighi et al., 2008; Schafer and Yarowsky, 2002) combined context information and other monolingual features (e.g., relative frequency and orthographic substrings,etc.) of source and target language words to learn translation pairs from monolingual corpora. Recently, several works (Mikolov, Le, and Sutskever, 2013a; Vulić and Moens, 2015; Vulić and Korhonen, 2016; Chandar et al., 2014; Wang et al., 2016) proposed cross-lingual word embedding strategies to map words from a source language vector space to a target language vector space, and also demonstrated its effective application to bilingual lexicon induction.

The approach presented here is similar to the end-to-end experiments of Irvine and Callison-Burch (2014) and Irvine and Callison-Burch (2016), but to generate bilingual lexica, instead of using a large variety of monolingual signals to learn and prune new phrase pairs, our method basically trained an SVM classifier using WE vector (Mikolov et al., 2013b), together with BC information (Brown et al., 1992) as features. To evaluate the impact of our blingual lexica on SMT, we conducted our experiment on Chinese (ZH)-Spanish (ES). Although they are two of the most widely spoken languages of the world, to the best of our knowledge, they are still suffering from the parallel data shortage problem. There are no direct SMT systems for this language pair but rule-based ones (Costa-Jussà and Centelles, 2014; Costa-Jussà and Centelles, 2016), which are still lacking a lot of coverage.

3 Approach

In this section, we describe a simple approach to improve the performance of a low resource SMT system by augmenting the phrase table with new translation pairs generated from

monolingual data. We treat bilingual lexicon generation as a binary classification problem: given a source word, the classifier predicts whether a target language word is its translation or not. Our classifier was trained with a seed lexicon of one thousand correct translation pairs (Section 4.1). We first used the concatenated WE of source and target word as features to train an SVM binary classifier following Han and Bel (2016). Then the trained model was used to find possible translations for a given source word among all target language vocabulary.

However, the first results showed that some words were wrongly considered as the translation of many different source words without being related to them in any meaningful way. This could be a consequence of the 'hubness problem' as reported by Radovanović et al. (2010). To improve the performance of our classifier, we decided to add BC representation to our WE features, since (Birch, Durrani, and Koehn, 2013; Matthews et al., 2014; Täckström, McDonald, and Uszkoreit, 2012; Agerri and Rigau, 2016) demonstrated that word clustering provides relevant information for cross-lingual tasks. Observing our data, semantically related words in the source monolingual corpus are grouped into the same class, while their translations belong to a corresponding class in the target monolingual corpus as well. For instance, in our ZH monolingual corpus, 演员(actor) and 记者(journalist) belong to the cluster 011111110110, while their translations *actor* and *periodista* are both grouped into the corresponding cluster 11010100 in the ES monolingual corpus. Therefore, we added BC of source and target words as additional features with the intention of helping the classifier to rule out those semantically unrelated target candidates to some extent.

To visualize the impact of using BC, in Figure 1, we plot the geometric arrangement of 6K word pairs (*right translation* and *no translation*[1]) represented by only WE vectors and with additional BC information in a 3-dimensional space. Each point represents a word pair since we concatenate the features of source and target words together. The change of the distribution of *right translation* or *no translation* demonstrates that the joint representation does encode relevant information for the classification.

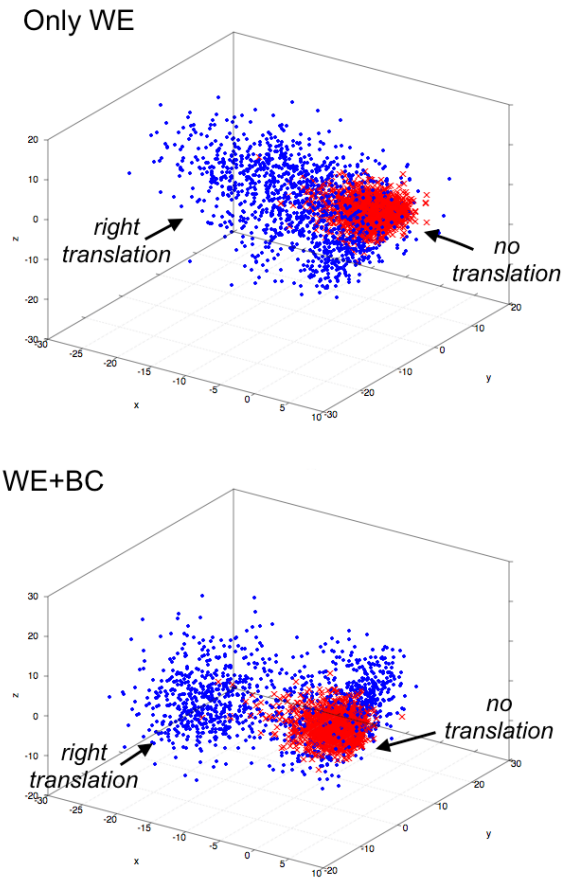

Figure 1: Distributed representations of 6K word pairs (1K *right translation* and 5K *no translation*) with WE of 400 dimensions and with combination of WE and BC of 800 dimensions. We used PCA to project high dimensional vector representations down into a 3-dimensional space

4 Experimental setup

In this section, we describe the experimental settings for evaluating our approach. The outline of our experiments is: (i) Generating the training positive and negative word pair lists. (ii) Obtaining the corresponding word embedding vector and (iii) Brown clusters from monolingual corpora. (iv) Concatenating the representation features of the source word and its translation equivalent (or random words for negative instances) (iv) Training an SVM classifier using the previously concatenated representations. (v) Producing new translation word pairs from monolingual corpora using the trained classifier. (vi) Training the SMT system with available par-

[1]The definition of *right translation* and *no translation* is given in section 4.1.

allel corpora plus the newly acquired translation word pairs.

4.1 Classifier datasets

To obtain the positive training set (*right translation*), a translation list was produced by first randomly extracting a list of about 1K nouns, verbs and adjectives[2] (frequency range from 10 to 100K) from the ZH monolingual corpus. Then these randomly selected words were translated from ZH to ES using on-line Google Translator and manually revised.

To build the negative training set (*no translation*), we randomly selected non-related words from the monolingual corpus of each language and randomly combined them. The ratio was 5 negative instances for each positive one[3]. The data set was split for training and testing: 1K positive and 5K negative word pairs for training; 300 positive and 1.5K negative word pairs for testing.

4.2 Word embedding

The monolingual corpora that were used for learning WE and BC were: Chinese Wikipedia Dump corpus[4] (149M words) and Spanish Wikipedia corpus[5] (130M words, 2006 dump). WE were created with the Continuous Bag-of-words (CBOW) method as implemented in the word2vec[6] tool, because it is faster and more suitable for large datasets (Mikolov, Le, and Sutskever, 2013a). To train the CBOW models we used the following parameters: window size 8, minimum word frequency 5 and 200 dimensions for both source and target vectors.

4.3 Brown clustering representation

Brown clusters[7] were induced from the same monolingual corpora that used for WE. We set c=200 for computational cost savings, although with larger number of clusters it might perform better. In order to include BC in word pair representations, instead of using directly the bit path, we used one-hot encoding. More specifically, 400 binary features were added to WE concatenated vectors: 200 for each word. Each component represents one of the 200 word clusters for each source and target word.

4.4 SVM Classifier

We built and tested an SVM[8] classifier on ZH-ES using the datasets described in Section 4.1 for three word categories: noun, adjective and verb. The evaluation was double, as we performed a 10 fold cross-validation with the training set and we tested again the model with a held-out test set.

4.5 Phrase-based SMT setup

Our SMT system was built using Moses phrase-based MT framework (Koehn et al., 2007). We used *mgiza* (Gao and Vogel, 2008) to align parallel corpora and *KenLM* (Heafield, 2011) to train a 3-gram language model. We applied standard phrase-based MT feature sets, including direct and inverse phrase and lexical translation probabilities. Reordering score was produced by a lexicalized reordering model (Koehn et al., 2005). The parameter 'Good Turing'[9] was applied in order to reduce overestimated translation probabilities, since the parallel corpus contained many unigram phrase pairs provided by our classifier. For the evaluation, we used BLEU metric (Papineni et al., 2002).

The parallel corpora that used to train and test the SMT system were: Chinese-Spanish OpenSubtitles2013[10] (1M sentences) for training; TAUS translation memory[11] (2K sentences) and UN corpus[12] (2K sentences) for testing. To train the language model, we combined Spanish Wikipedia corpus mentioned in Section 4.2 with OpenSubtitles2013 target corpus.

The classifier was used to deliver, for each of about 3K selected source words (the most frequent words that were not present in the

[2] For PoS tagging of all corpora, we used the Stanford PoS Tagger (Toutanova, Dan Klein, and Singer, 2003).

[3] We chose this unbalanced ratio to approach the actual distribution of the data to classify since there will be many more *no translation* than *right translation* pairs.

[4] https://archive.org/details/zhwiki_20100610

[5] http://hdl.handle.net/10230/20047

[6] https://code.google.com/archive/p/word2vec/

[7] https://github.com/percyliang/brown-cluster

[8] As implemented in WEKA (Hall et al., 2009).

[9] http://www.statmt.org/moses/?n=FactoredTraining.ScorePhrases

[10] http://opus.lingfil.uu.se/OpenSubtitles2013.php

[11] http://www.tauslabs.com/

[12] http://opus.lingfil.uu.se/UN.php

baseline phrase table), all the possible translation candidates as found in the combination with the 30K target words of the same PoS (for computational saving). All word pairs classified as right translation were then appended to the existing parallel corpora for training a new SMT system. Figure 2 shows the generation and integration of the new translation pairs.

Input: Vector representations of 3K source words *S1*; Vector representation of all target words *T1*; Supervised classifier model *M*; Parallel corpora for SMT baseline *L1*
Output: Expanded parallel corpora *L2*
for each source word vector *V*(*x*) in *S1* **do**
  for each target word vector *V*(*y*) in *T1* **do**
    if PoS of source word *x* and target word *y* are the same **then**
      concatenate *V*(*x*) with *V*(*y*);
      append the concatenation *V*(*x*,*y*) to *C*;
    end
  end
end

for each concatenation *V*(*x*, *y*) in *C* **do**
  test *V*(*x*, *y*) using *M*;
  if *V*(*x*, *y*) is classified as 'right translation' **then**
    append the word pair (*x*, *y*) to *L1*;
  else
    pass
  end
end

Figure 2: Algorithm for the generation and integration of supervised bilingual lexicons

5 *Experimental results*

We present here the evaluation results of the classifier and their impact on our low resource ZH-ES SMT system.

5.1 Results on bilingual lexicon induction

Table 1 shows the evaluation results of our classifier trained with WE and with the combination of WE and BC in terms of precision (P), recall (R) and F1-measure (F).

Evaluation results show that the classifiers are capable of finding out the correct translation among all the candidates with same PoS in the target monolingual corpus in most of the cases. With the classifier trained only using WE, we already obtained a precision and recall of 0.926 and 0.87, respectively for *right translation*. To explore the relation between

		10 cross-validation			Held-out test set		
		P	R	F1	P	R	F1
WE	Yes	0.937	0.919	0.928	0.926	0.87	0.89
	No	0.984	0.988	0.986	0.976	0.987	0.981
WE+BC	Yes	0.955	0.935	0.945	0.955	0.92	0.937
	No	0.987	0.991	0.989	0.985	0.992	0.988

Table 1: Test results of the ZH-ES classifier trained with WE and with WE+BC

the performance of the classifier and the number of training instances, Figure 3 plots the learning curves (F1, and kappa value) over different percentages of positive training instances from 100 (10%) to 900 (90%), with corresponding negative instances from 500 to 4500. It shows that the classifier achieved stable and good results with around 50% of the training instances.

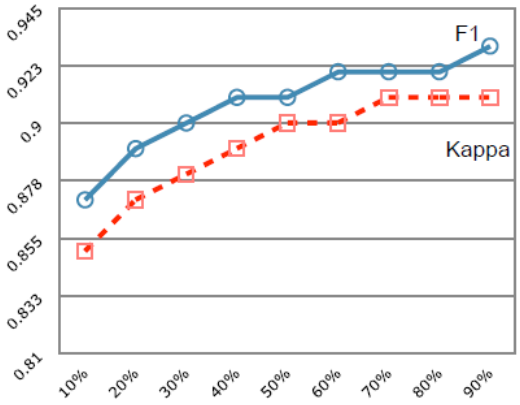

Figure 3: Learning curve over different percentages of the training data for Chinese and Spanish

However, the classifier trained using only WE was not efficient in the following cases:

(i) Candidates affected by *hubness problem*.

After an error analysis, we realized that a small group of target words were repeatedly assigned as possible translations to many different source words, such as *parte* ('part'), *nombre* ('name') and *tiempo* ('time').

(ii) Semantically related candidates.

Words that always occur in similar contexts or nearby tended to be confusing for the classifier to make the right decision. For instance, the classifier assigned both turista ('tourist'), turismo ('tourism') as possible translations for the source word 旅游业('tourism').

After adding BC, both precision and recall results were improved as shown in Table 1, demonstrating that BC indeed provided rel-

evant information for ruling out many wrong translation candidates. In terms of accuracy, with BC the performance improved from 96.8 to 97.6, resulting in a considerable reduction of the number of word pairs classified as right translation. Note that 88.65 M word pairs were presented to the classifier from 2955 source words combined with 30K target words. The WE classifier delivered a 7% word pairs classified as *right translation*, while the WE+BC classifier delivered only a 2.7%.

In order to verify whether the classifier was not learning that particular BCs were associated to the right or wrong translation categories, we checked the distribution of the clusters in both categories: 57 different clusters were present in both positive and negative examples in the training data set and 23 in the test set.

5.2 Evaluation on SMT translation table expansion

Table 2 shows experimental results of the SMT system trained using the enriched parallel corpora. The system was tested on two different test sets (described in 4.5)and measured by BLEU metric and Out of Vocabulary rate (OOV).

Setup	TAUS		UN	
	BLEU	OOV	BLEU	OOV
Baseline	8.8	9.6%	10.81	6.8%
Baseline + 3K SBL	9.58	8.7%	11.42	5.9%

Table 2: BLEU and OOV test results of the baseline and the system developed with our supervised bilingual lexica (SBL)

According to the results shown in Table 2, with the new translation candidates given by our classifier, the performance of the SMT system improved with respect to the baseline by up to +0.70 and +0.61 BLEU scores, and the OOV[13] rate of baseline system was reduced around 0.9% for both test sets.

Table 3 shows several examples of translation outputs after adding the bilingual lexica compared to the results of the baseline SMT system. Note that although all possible translation candidates delivered by the classifier are included, the SMT system is able to find out the right translation, thus improving the quality of the translations with respect to OOV, as expected.

Source: 文化多样性(Cultural diversity) **Reference**: diversidad cultural **Baseline**: 多样性. a la cultura **Baseline+SBL**: diversidad cultural
Source: 负面影响(Negative impact) **Reference**: consecuencias negativas **Baseline**: la negativo **Baseline+SBL**: un impacto negativo
Source: 继续支助(continue supporting) **Reference**: continúen apoyando **Baseline**: seguir 支助 **Baseline+SBL**: estado manteniendo

Table 3: Translation examples of our SMT baseline and the system with acquired lexicons

6 Conclusions

This paper described a supervised approach to automatically learn bilingual lexicons from monolingual corpora for improving the performance of a Chinese to Spanish SMT system. Our experiment shows an improvement of +0.7 BLEU score is achieved even though an average of 800 translation pairs per source word were added to the existing parallel corpus. The high recall of our classifier ensures that more reliable translation candidates can be introduced to the SMT system and the language model component is able to handle the selection of the correct one, hence delivering a better translation output. To further improve the performance of the classification, our future work includes combining our model with other models separately trained on multiple monolingual features using ensemble learning.

7 Acknowledgments

Han Jingyi was supported by the FI-DGR grant program of Generalitat de Catalunya.

References

Agerri, R. and G. Rigau. 2016. Robust multilingual named entity recognition with shallow semi-supervised features. *Artificial Intelligence*, pages 63–82.

Birch, A., N. Durrani, and P. Koehn. 2013. Edinburgh slt and mt system description for the iwslt 2013 evaluation. *in Proceedings of the 10th International Workshop*

[13]The OOV words were generated as shown in:`http://www.statmt.org/moses/?n=Advanced.OOVs`

on Spoken Language Translation, pages 40–48.

Brown, P. F., P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, pages 467–479.

Chandar, S., S. Lauly, H. Larochelle, M. Khapra, B. Ravindran, V. C. Raykar, and A. Saha. 2014. An autoencoder approach to learning bilingual word representations. *Advances in Neural Information Processing Systems*, pages 1853–1861.

Chiao, Y.-C. and P. Zweigenbaum. 2002. Looking for candidate translational equivalents in specialized, comparable corpora. *in Proceedings of the 19th International Conference on Computational Linguistics*, pages 1208–1212.

Costa-Jussà, M. R. and J. Centelles. 2014. Chinese-to-spanish rule-based machine translation system. *in Proceedings of the EACL Workshop on Hybrid Approaches to Translation (HyTra)*.

Costa-Jussà, M. R. and J. Centelles. 2016. Description of the chinese-to-spanish rule-based machine translation system developed using a hybrid combination of human annotation and statistical techniques. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Fung, P. 1995. Compiling bilingual lexicon entries from a non-parallel english-chinese corpus. *in Proceedings of the Third Workshop on Very Large Corpora*, pages 173–183.

Gao, Q. and S. Vogel. 2008. Parallel implementations of word alignment tool. *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 49–57.

Haghighi, A., P. Liang, T. Berg-Kirkpatrick, and D. Klein. 2008. Learning bilingual lexicons from monolingual corpora. *in Proceedings of the annual meeting on Association for Computational Linguistics*, pages 771–779.

Han, J. and N. Bel. 2016. Towards producing bilingual lexica from monolingual corpora. *in Proceedings of the International Language Resources and Evaluation*, pages 2222—2227.

Heafield, K. 2011. Kenlm: Faster and smaller language model queries. *in Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197.

Irvine, A. and C. Callison-Burch. 2013. Supervised bilingual lexicon induction with multiple monolingual signals. *in Proceedings of HLT-NAACL '13*, pages 518–523.

Irvine, A. and C. Callison-Burch. 2014. Hallucinating phrase translations for low resource mt. *in Proceedings of the Conference on Computational Natural Language Learning*, pages 160–170.

Irvine, A. and C. Callison-Burch. 2016. End-to-end statistical machine translation with zero or small parallel texts. *Natural Language Engineering*, pages 517–548.

Koehn, P., A. Axelrod, A. B. Mayne, C. Callison-Burch, M. Osborne, and D. Talbot. 2005. Edinburgh system description for the 2005 iwslt speech translation evaluation. *MT summit*, pages 79–86.

Koehn, P., A. B. Hieu Hoang, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open source toolkit for statistical machine translation. *in Proceedings of the annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180.

Koehn, P. and K. Knight. 2002. Learning a translation lexicon from monolingual corpora. *ACL Workshop on Unsupervised Lexical Acquisition*, pages 9–16.

Marton, Y., C. Callison-Burch, and P. Resnik. 2009. Improved statistical machine translation using monolingually-derived paraphrases. *in Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 381–390.

Matthews, A., A. Waleed, A. Bhatia, W. Feely, G. Hanneman, E. Schlinger, S. Swayamdipta, Y. Tsvetkov, A. Lavie, and C. Dyer. 2014. The cmu machine

translation systems at wmt 2014. *in Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 142–149.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013b. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Q. V. Le, and I. Sutskever. 2013a. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

Mirkin, S., N. C. Lucia Specia, I. Dagan, M. Dymetman, and I. Szpektor. 2009. Source-language entailment modeling for translating unknown terms. *in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 791–799.

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. *in Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Radovanović, M., A. Nanopoulos, and M. Ivanović. 2010. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research 11*.

Rapp, R. 1995. Identifying word translations in non-parallel texts. *in Proceedings of the annual meeting on Association for Computational Linguistics*, pages 320–322.

Razmara, M., M. Siahbani, G. Haffari, and A. Sarkar. 2013. Graph propagation for paraphrasing out-of-vocabulary words in statistical machine translation. *in Proceedings of the annual meeting on Association for Computational Linguistics*.

Saluja, A., H. Hassan, K. Toutanova, and C. Quirk. 2014. Graph-based semi-supervised learning of translation models from monolingual data. *in Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 676–686.

Schafer, C. and D. Yarowsky. 2002. Inducing translation lexicons via diverse similarity measures and bridge languages. *in Proceedings of the Conference on Natural Language Learning*, pages 1–7.

Toutanova, K., C. M. Dan Klein, and Y. Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. *in Proceedings of HLT-NAACL'03*, pages 252–259.

Turchi, M. and M. Ehrmann. 2011. Knowledge expansion of a statistical machine translation system using morphological resources. *Research Journal on Computer Science and Computer Engineering with Application (Polibits)*.

Täckström, O., R. McDonald, and J. Uszkoreit. 2012.

Vulić, I. and A. Korhonen. 2016. On the role of seed lexicons in learning bilingual word embeddings. *in Proceedings of the annual meeting on Association for Computational Linguistics*, pages 247–257.

Vulić, I. and M.-F. Moens. 2015. Bilingual word embeddings from non-parallel document-aligned data applied to bilingual lexicon induction. *in Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 719–725.

Wang, R., H. Zhao, S. Ploux, B.-L. Lu, M. Utiyama, and E. Sumita. 2016. A novel bilingual word embedding method for lexical translation using bilingual sense clique. *arXiv preprint arXiv:1607.08692*.

Yu, K. and J. Tsujii. 2009. Bilingual dictionary extraction from wikipedia. *in Proceedings of the twelfth Machine Translation Summit*, pages 379—386.